

卒業論文 2016 年度（平成 28 年度）

期待和了平均順目の評価による  
モンテカルロ木探索の麻雀への適用

慶應義塾大学 環境情報学部

細田 航星

徳田・村井・楠本・中村・高汐・バンミーター・植原・三次・中澤・  
武田 合同研究プロジェクト

2017 年 1 月



卒業論文 2016 年度（平成 28 年度）

## 期待和了平均順目の評価による モンテカルロ木探索の麻雀への適用

### 論文要旨

近年、2 人零和確定完全情報ゲームであるチェス、オセロ、将棋といったゲームでは人間のトッププレイヤーと同等かそれ以上の実力を持つコンピュータプレイヤーが提案されている。一方で多人数不完全情報ゲームである麻雀では未だそれに及ばない状態である。これは従来のゲーム木の探索手法を適用するのが難しいためである。本論文では、探索空間が大きくなることでうまく適用できなかったモンテカルロ法について、期待和了平均順目の数理的な評価を利用することで探索空間を小さくし、適用する手法を提案する。この手法の効果を確認する実験として、多人数性を排除した 1 人麻雀における和了率の比較を、人間のプレイヤーや先行研究 AI と行った。またそれをそのまま 4 人麻雀で対戦させた場合の比較も行う。解析の結果、本手法で構築した AI は 1 人麻雀において、平均レベルのプレイヤーの和了率を出すことができることがわかった。また、4 人麻雀に適用した場合、シャンテン数を下げるように打つ AI に比べて高いレートを出すことが可能であることがわかった。

### キーワード

麻雀, モンテカルロ法, ゲーム情報学

慶應義塾大学 環境情報学部

細田 航星



# **Abstract Of Bachelor's Thesis    Academic Year 2016**

## **Summary**

In recent years, computer games that are equal to or better than the top players of humans have been proposed for games such as chess, Othello, and shogi, which are two-person zero settled perfect information games. On the other hand, Mahjong which is a multiplayer incomplete information game has not yet reached it. This is because it is difficult to apply conventional search methods of game trees. In this paper, we propose a method to apply by restricting mathematical evaluation and applied parts of the Monte Carlo method which could not be applied successfully by increasing the search space. For evaluation, comparison of the rate of winning in one mahjong excluding multiplayer was done with human players and previous studies AI. We also make comparisons when we played it with 4 mahjong as it is. As a result of the analysis, we found that the AI constructed by this method has comparable ability to that of the average level player.

## **Keywords**

Bachelor of Arts in Environmental Information  
Keio University

Wataru Hosoda



# 目次

第1章	序論	1
1.1	本論文の背景	1
1.2	本論文が着目する課題	1
1.3	本論文の目的	2
1.4	本論文の構成	2
第2章	先行研究	3
2.1	人間プレイヤーの牌譜から学習する手法を用いた研究	3
2.2	麻雀についてモンテカルロ法を応用した研究	3
2.2.1	UCB1	3
2.2.2	Lin UCB	4
第3章	提案手法	5
3.1	1人麻雀について	5
3.2	探索空間を限定する手法の提案	6
3.2.1	1人麻雀プレイヤーにおける和了率の数理的評価	6
3.2.2	二向聴以下の牌姿に対してモンテカルロ法を適用する	6
第4章	評価	9
4.1	実験概要	9
4.2	1人麻雀プレイヤーによる和了率の比較	9
4.3	4人麻雀における先行研究とのレートと比較	9
第5章	結論	11
5.1	本研究のまとめ	11
5.2	今後の課題と展望	11
	参考文献	12

## 图 目 次



## 表 目 次



# 第1章 序論

## 1.1 本論文の背景

近年、2人零和確定完全情報ゲームであるチェス、オセロ、将棋といったゲームではAIの研究がめざましく、人間のトッププレイヤーと同等以上の成績を記録している。一方で多人数不完全情報ゲームである麻雀では未だそれに及ばない状態である。これは従来のゲーム木の探索手法を適用するのが難しいためである。多人数ゲームでは状況によってプレイヤーの行動の目的が違ふといった難しさがある。この難しさを解消するために、先行研究 [10][9] でも多数行われている通り、相手を考慮しない1人麻雀プレイヤーに対して和了率を求める評価を採用する。麻雀を行う時の人間の思考は1人麻雀をもとにしているため、1人麻雀から4人麻雀への拡張は容易だと考えられる。

## 1.2 本論文が着目する課題

1人麻雀においてある局面からゲーム木を展開しようとした場合、次にどの種類牌をツモるかはわからないため、ランダムに決定しノードを展開していくことになる。これを繰り返す行くと探索空間が大きくなりすぎるため、手の選択が難しくなる。これを対処するために、あらゆる手法が提案されている。このような場合に木を展開せずに有望な手を選択する手法として、Upper Confidence Bound (UCB) [1] がある。UCB ではある局面から考えられる全ての手に対して、よい結果を返しそうな手を重視しつつ、何度もプレイアウトと呼ばれるランダムシミュレーションを行うことで最善の手を決定する手法である。UCB は局面ごとにこの探索を実行するが、各局面を完全に別の局面として扱うため他の局面の探索において得た情報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表すことで、対象とする局面が異なっても、それまで対象とした局面の情報を利用できるといったものである。これらは麻雀について適用した例が報告されている [10] が、いずれの方法も平均プレイヤーに満たない成績となった。

本研究では、モンテカルロ法の問題点である、麻雀に適用すると探索空間が大きくなりすぎてうまく適用出来ない問題に対して、探索空間を小さくするために、2つの解決策を行う。一つはモンテカルロ法のシミュレーションを直接牌姿に当てはめるのではなく、牌姿ごとの局面で数理的に比較できる内容を数理計算を用い削減する方法である。もう一つは、三向聴以下に対してはシャンテン数を下げるように打ち、モンテカルロ法を用いる部分を限定する方法である。麻雀の牌理に関しては、上がりに近づく選択ほど重要である [11] ため、聴

牌に近い打牌選択においてモンテカルロ法を適用することで、成績に影響を大きく与える部分でモンテカルロ法の効果が発揮できることが期待される。

### 1.3 本論文の目的

本研究では和了率を理論値計算によって評価することと和了に近い段階へのモンテカルロ法の部分適用によって、麻雀に対してモンテカルロ法をそのままうまく適用できない問題に対して解決策を提案する。多人数ゲームでは状況によってプレイヤーの行動の目的が違ふといった難しさがある。この難しさを解消するために、先行研究 [1,2] でも多数行われている通り、相手を考慮しない 1 人麻雀プレイヤーに対して和了率を求める評価を採用する。従来のモンテカルロ法を麻雀に適用した例では平均プレイヤーレベルの実力が出ていないため、本手法によってそれらを上回ることが期待される。

### 1.4 本論文の構成

本論文の構成について述べる。2 章では、不完全情報ゲームの関連研究について述べる。3 章では、1 人麻雀にモンテカルロ法を適用する際に、適用範囲を限定し、比較内容を数理的和了率によって評価する手法を提案する。4 章では、3 章で提案した手法で実装した AI に 1 人麻雀を打たせ、各種パラメータを比較する。また、4 人麻雀で打った際の成績の評価も行う。5 章では、本手法によって得られた知見について述べる。

## 第2章 先行研究

本章では不完全情報ゲームの関連研究について述べる。麻雀の研究としては以下の研究が報告されている。

### 2.1 人間プレイヤーの牌譜から学習する手法を用いた研究

人間プレイヤーの牌譜を教師信号とした麻雀評価関数の機械学習の報告がいくつかなされている。北川らは評価関数に3層ニューラルネットワークを用いた教師あり学習を用いることで麻雀AIのパラメータ調整を行った。[7] 牌譜とAIの一致率はツモ局面において約56%, 鳴き局面において約89%, 東風荘で得られたレートは1318であった[6]。三木らは木カーネルを用いた非線形SVMによって手牌の分類を学習した。ツモ局面における人間プレイヤーとの一致率は51%であった[5]。しかし、いずれの方法も人間の平均プレイヤーの実力に達していない。

水上ら[?] 1人麻雀プレイヤーの学習に牌譜との一致を目指した平均化パーセプトロンを用いた。麻雀の特徴量は非常に膨大なため、この方法が採用された。学習に使った教師信号は、麻雀サイト天鳳[6]において鳳凰卓でプレイすることができるプレイヤーの試合データである。鳳凰卓でプレイできるのは全プレイヤーの中でも上位0.1%程度であり牌譜の質は高いと考えられる。この研究では、1人麻雀プレイヤーの和了率が平均プレイヤーを上回り、鳳凰卓でプレイできるプレイヤーの和了率に近いレベルになった。しかし、1人麻雀プレイヤーの学習を行う際に、4人麻雀の牌譜を使っていることで、4人麻雀独特の打ち方を排除できない問題や、上級者が認知できない最適解を求められない問題があった。

### 2.2 麻雀についてモンテカルロ法を応用した研究

#### 2.2.1 UCB1

UCB1はUCB1値が最大となるノードに対してプレイアウトを行い、その結果によりUCB1値を更新するという手順を一定回数繰り返し、最も平均報酬が高い選択枝を選ぶアルゴリズムである。UCB1値は、 $j$ は子ノード $j$ の平均報酬、 $\alpha$ は定数、 $n$ は親ノードの探索回数、 $n_j$ は子ノード $j$ とした時、 $UCB = x_j + \alpha \sqrt{2 \log n / n_j}$  (1) で表される。 $x$ 式(1)右辺の第1項は平均報酬を、第2項は信頼度を示している。信頼度はそのノードのプレイアウト回数が少ないと大きく、多いと小さくなる。UCB1値を用いることでUCB1ではより

有望そうな手に対して多くのシミュレーションを行う事ができる。UCB1 は各局面ごとにこの手順を実行するが、各局面を完全に異なる局面として扱うため、他の局面の探索において得られた情報を共有することができないという欠点を持つ。この欠点を補うアルゴリズムとして期待されているのが LinUCB である。

### 2.2.2 Lin UCB

LinUCB [5] は、UCB を局面を特徴で表すことができるように拡張したものであり、牌譜の局面からの教師あり学習や異なる探索の結果の共有ができる手法である。

## 第3章 提案手法

本研究での提案手法は麻雀の多人数を削除した、1 人麻雀を考え、その和了率の最大化を図る手法を提案する。

### 3.1 1 人麻雀について

多人数ゲームでは状況によってプレイヤーの行動の目的が違うといった難しさがある。この難しさを解消するために、相手を考慮しない 1 人麻雀プレイヤーを考える。この 1 人麻雀プレイヤーのルールは、相手プレイヤーを考えないため、相手プレイヤーに点数を支払う事による放銃や被ツモ失点などを考えない。また、相手プレイヤーによる捨て牌が存在しないので、鳴きや栄和を考えない。リーチについても、和了率の観点では不要なため、考えない。門前でツモを繰り返し行うだけのシンプルなシステムである。先行研究 [?] 評価を合わせるため、ツモの回数は 27 回とした。

プレイ人数に着目した研究としてはポーカーを用いたもの [3],[4] がある。これらの手法では 2 プレイヤでのゲームを強くしてから多人数に適応する方法がとられている。ポーカーは 2 プレイヤであればナッシュ均衡戦略を用いることで世界チャンピオンに勝っているが、ポーカーは 2 人から 10 人程度で参加可能なゲームであり人数が増えると状態数が指数関数的に増大するためナッシュ均衡戦略の計算は難しい。そこで 2 人で行われたナッシュ均衡戦略を 3 人限定で拡張する方法、プレイヤーの行動を削減、抽象化することでより少ない人数の少ないゲームを想定する方法がとられた。

また、これらを元に麻雀においてもプレイヤーの人数を削減してから多人数に適用する方法が取られている。水上ら [?] 1 人麻雀プレイヤーの学習に牌譜との一致を目指した平均化パーセプトロンを用いた。麻雀の特徴量は非常に膨大なため、この方法が採用された。学習に使った教師信号は、麻雀サイト天鳳 [6] において鳳凰卓でプレイすることができるプレイヤーの試合データである。鳳凰卓でプレイできるのは全プレイヤーの中でも上位 0.1 %程度であり牌譜の質は高いと考えられる。この研究では、1 人麻雀プレイヤーの和了率が平均プレイヤーを上回り、鳳凰卓でプレイできるプレイヤーの和了率に近いレベルになった。しかし、1 人麻雀プレイヤーの学習を行う際に、4 人麻雀の牌譜を使っていることで、4 人麻雀独特の打ち方を排除できない問題や、上級者が認知できない最適解を求められない問題があった。

1 人麻雀において和了率を最大化することに関しては、以下のような理由が述べられる。

4 人麻雀では平均順位の低い (つまり上手い) プレイヤほど、平均和了点が低くても和了率が高いことが統計で明らかになっている [8]。上手いプレイヤーの平均和了点が低くなる理

由として手役を無理に狙った打ち方をしないためである。また麻雀の点数の特性上、満貫以上は難易度の割には点数が高くないといったことが挙げられるため、比較的点数の低い点数で多く和了することが効率が良い。他にも和了すれば他プレイヤーが和了することができないため、相手のチャンスをつぶすといった意味でも和了率の高さは大事である。

## 3.2 探索空間を限定する手法の提案

### 3.2.1 1人麻雀プレイヤーにおける和了率の数理的評価

国土氏の研究により、ある手牌における聴牌までの平均消費巡目は、それぞれの変化での平均消費巡目のそれぞれの変化する確率での単純な平均で与えられることがわかっている。[3] すなわち、元の成功率が  $p$ 、手変わりする確率が  $q, r$ 、手変わり後の成功率が  $Q, R$  ならば、

その手の向聴が進むまでの平均消費巡目は

$$(p * 1/p + q * 1/Q + r * 1/R) / (p + q + r) \quad (2)$$

と表される。

本論文ではこれを和了時までの式に拡張する。まず、聴牌までの平均順目と聴牌後の和了までの平均順目の合計がすなわち全体の平均消費順目であることを証明する。次に、それを合成した結果の数式がどうなるかを書く。

考察の結果このようになる。

$$(p * 1/p + q * 1/Q + r * 1/R) / (p + q + r) + \text{和了期待順目} \quad (3)$$

また、(現在の平均テンパイ巡目)=(次巡の平均テンパイ巡目)+1 であることから、各ノードが手替わり率とテンパイ率を値として持つ深さ 1 の木構造で記述できる。したがって、同様に、 $n$  手先の手替わりまで考えると深さ  $n$  の木構造で記述することができる。 $n$  を大きくすることにより、精度の高い和了率を近似することが可能となる。

図 2

本研究では、この式 (3) をそれぞれの牌姿に当てはめ、その計算結果が高くなるようなノードを探索する。これにより、従来の手法では何回も同じ牌姿の変化をシミュレートしなければいけない問題があったが、(3) 式の評価によってその多くが削減できる。したがって、この方法によって精度が高くなることが期待される。

### 3.2.2 二向聴以下の牌姿に対してモンテカルロ法を適用する

麻雀において、和了までの手順の中では、シャンテン数が小さい時点での選択が成績に影響を与えやすいことがわかっている [11] したがって、成績向上のためには和了系に近い部分において戦略を改善することが成績に影響を与えやすい事がわかる。

また、モンテカルロ法の問題点は、前述したとおり麻雀に適用すると探索空間が大きくなりすぎることが問題であった。しかしこれを和了系に近い部分に適用することで、探索空間



を小さく削減することができるため、適用することができるようになる。本研究の手法である与えられた牌姿における和了率の近似式（3）を適用する際にも、シャンテン数が大きい場合については正確に見積もることが難しいため、このような限定的な部分への適用が重要である。

本論文では二向聴以上の部分にこのモンテカルロ法を適用することで、従来の問題点を解決する。



## 第4章 評価

### 4.1 実験概要

### 4.2 1人麻雀プレイヤーによる和了率の比較

提案手法で提示した、期待和了平均順目の評価によるモンテカルロ木探索を1人麻雀に適用し、性能を評価した。期待和了平均順目を $n$ 順先の変化まで評価することによって性能が変わる可能性があるので、1順、2順、3順までの変化を考慮するものをそれぞれ分けて実験した。それぞれに対して100局のテストデータと10,000局のテストデータを与え、あがることができた局数を計測した。テストデータとは、全ての牌をランダムに並べたものを100セット用意したデータである。このデータから13牌を初期牌として与え、その後牌を引いて切る動作を27回行い、その中であがれたかどうかを確認した。100局のテストデータで人間のプレイヤーとの比較を行い、10,000局のテストデータで各手法の性能の評価を行った。

### 4.3 4人麻雀における先行研究とのレートとの比較

本研究では1人麻雀における和了率の向上を目指すため、和了率の数理的評価とモンテカルロ法を適用する部分を限定する手法をとった。佐藤らは、1人麻雀における和了率を有効牌を数え上げて大きくなるようにすることで、和了率の最大化を図り、これを4人麻雀で打たせレートを取った。同じように本研究手法でも4人麻雀で打たせた結果を比較した。

表

このように、レートは本手法の方が上回る結果となった。ここで保証安定レートについて比較をする。



## 第5章 結論

### 5.1 本研究のまとめ

本研究では、各牌姿における期待和了平均順目の評価をモンテカルロ木探索に利用することによって、モンテカルロ木探索を麻雀に適用することを提案した。評価として、1人麻雀の和了率を比較した。先行研究ではUCB 1やLinUCBを1人麻雀に適用する例が報告されていたが、これらは和了までのシミュレーションを行っているものであり、探索空間が多く精度が落ちるため平均プレイヤには及ばなかった。本研究の提案手法では、和了率が平均プレイヤ達することが確認できた。この結果から、麻雀において従来の探索方法では探索空間が大きくなりすぎて精度が落ちる問題に対して、期待平均和了順目を用いて和了への確率を簡潔に評価することにより、探索空間を小さくする方法が有用であることがわかった。また、期待平均和了順目をどれほど深くの変化まで計算するかに関しては、深く計算することでより精度が保たれることが確認できた。しかし、5巡目以降の変化について考慮した場合とそうでない場合において大きな違いが見られないため、期待平均和了順目はおよそ5巡目までの変化を考慮することで十分な精度が期待できることがわかった。4人麻雀の評価については、シャンテン数を下げるように打った先行研究よりもレートが高いということを示すことが出来た。したがってこの結果から、モンテカルロ木探索を期待平均和了巡目を指標に行う手法については有用だと考えることができる。

### 5.2 今後の課題と展望

今回の検証では、既存のモンテカルロ木探索を行った研究よりも、1人麻雀の和了率は高くなることを示すことが出来た。しかし、牌譜から教師信号を学習した方法により和了率を上回することは出来ていない。この原因としては、期待平均和了順目を用いる方法では、ツモを無限思考回数行える場合の平均順目を最小にするように考えていることである。実際の麻雀では、ツモ回数が限られていて、相手の和了によってもその終局までの巡目が異なる。今後の課題としては、期待平均和了順目に対して、残りツモ回数の変数を導入したモデルを構築し、同じようにモンテカルロ木探索を行うことである。このようにすることで、より4人麻雀に近い状態での和了率最大化を目指すことができる。



## 参考文献

- [1] Auer, p., cesa-bianchi, n. and fischer, p.: Finite-time analysis of the multiarmed bandit problem, machine learning, vol. 47, no. 2-3, pp. 235256 (2002).
- [2] Li, l., chu, w., langford, j. and schapire, r. e.: A contextual-bandit approach to personalized news article recommendation, proceedings of the 19th international conference on world wide web, acm, pp. 661670 (2010).
- [3] Risk, n. a. and szafron, d.: Using counterfactual regret minimization to create competitive multiplayer poker agents, proceedings of the 9th international conference on autonomous agents and multiagent systems: volume 1-volume 1, international foundation for autonomous agents and multiagent systems, pp. 159166 (2010).
- [4] 古居敬大:相手の抽象化による多人数ポーカーの戦略の決定, 修士論文, 東京大学 (2013).
- [5] 三木理斗:多人数不完全情報ゲームにおける最適行動決定に関する研究, 修士論文, 東京大学 (2010).
- [6] 天鳳:<http://tenhou.net/> (2013).
- [7] 北川竜平:麻雀の牌譜からの打ち手評価関数の学習, proceedings of 12th game programming workshop (2007).
- [8] とつげき東北. おしえて!科学する麻雀, 2009.
- [9] 水上 直紀ほか. 降りるべき局面の認識による 1 人麻雀プレイヤの 4 人麻雀への適用. 第 18 回ゲームプログラミングワークショップ (GPW2013) , Vol. 31, No. 3, pp. 1-7, 2013.
- [10] 中張 遼太郎ほか. Linucb の 1 人麻雀への適用. ゲームプログラミングワークショップ, 2013.
- [11] ネマタ. テンパイの技術. 勝つための現代麻雀技術論, 2014.